

Tilburg University

Perception of intersensory synchrony

Vroomen, J.; Keetels, M.N.

Published in:
Attention, Perception & Psychophysics

DOI:
[10.3758/APP.72.4.871](https://doi.org/10.3758/APP.72.4.871)

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Vroomen, J., & Keetels, M. N. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception & Psychophysics*, 72(4), 871-884. <https://doi.org/10.3758/APP.72.4.871>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

TUTORIAL REVIEW

Perception of intersensory synchrony: A tutorial review

JEAN VROOMEN AND MIRJAM KEETELS

Tilburg University, Tilburg, The Netherlands

For most multisensory events, observers perceive synchrony among the various senses (vision, audition, touch), despite the naturally occurring lags in arrival and processing times of the different information streams. A substantial amount of research has examined how the brain accomplishes this. In the present article, we review several key issues about intersensory timing, and we identify four mechanisms of how intersensory lags might be dealt with: by ignoring lags up to some point (a wide window of temporal integration), by compensating for predictable variability, by adjusting the point of perceived synchrony on the longer term, and by shifting one stream directly toward the other.

Because many natural events can be perceived via multiple senses, we typically have access to multiple features of those events across the different senses. For example, a speaker can be heard and seen at the same time. Audio–visual speech, however, is one example of the multisensory nature of perception, and there are many others. To take another, in order to decide whether a visual object is moving or stable, one needs to combine the visual information from the retina with kinesthetic and motor information about any motion of the viewer’s eyes, head, or entire body. Perception, even if called “visual” or “auditory,” is thus, in essence, multisensory, a point made long ago by Gibson (1966). The multisensory nature of the world is highly advantageous, because it increases perceptual reliability and saliency, and, as a result, it can enhance learning, discrimination, or the speed of a reaction to the stimulus (e.g., Sumbly & Pollack, 1954; Summerfield, 1987). However, the multisensory nature of perception also raises the question about how the sense organs cooperate so as to form a coherent representation of the world. In recent years, this multisensory nature of perception has been the focus of much behavioral and neuroscientific research (Calvert, Spence, & Stein, 2004). The most commonly held view among researchers in multisensory perception is what has been referred to as the *assumption of unity*. It states that, as information from different modalities share more (amodal) properties, the more likely it is that the brain treats them as originating from a common object or source (see, e.g., Bedford, 1989; Bertelson, 1999; Radeau, 1994; Spence, 2007; Stein & Meredith, 1993; Welch, 1999; Welch & Warren, 1980). Without doubt, the most important amodal property is *temporal coincidence*

(e.g., Radeau, 1994). From this perspective, one expects intersensory interactions to occur if, and only if, the information from the different sense organs reaches the brain at around the same time; otherwise, separate events are perceived, rather than a single multisensory one.

The perception of time, however, and, in particular, synchrony among the senses, is not straightforward, because no sense organ registers time on an absolute scale. Moreover, to perceive synchrony, the brain must deal with differences in physical and neural transmission times. Sounds, for example, travel through air much more slowly than does light (330 vs. 300,000,000 m/sec), whereas no physical transmission time through air is involved for tactile stimulation, which is usually presented directly at the body surface. The neural processing time also differs among the senses, being typically slower for visual stimuli than for auditory stimuli (approximately 50 vs. 10 msec, respectively), whereas, for touch, the brain may have to take into account where the stimulation originated, because the traveling time is longer from the toes to the brain than from the nose (the typical conduction velocity is 55 m/sec, which results in a ~30-msec difference between toe and nose for a distance of 1.60 m; Macefield, Gandevia, & Burke, 1989). Because of these physical and neural differences, it has been argued that auditory and visual information arrives synchronously at the primary sensory cortices only if the event occurs at a distance of approximately 10–15 m from the observer. This has been called the *horizon of simultaneity* (Pöppel, 1985; Pöppel, Schill, & von Steinbüchel, 1990), assuming that, arguably, synchrony is perceived at the primary sensory cortices. Sounds should thus appear to arrive before visual stimuli if the audio–visual event is within 15 m of

the observer, whereas vision should arrive before sounds for events farther away. Surprisingly, however, despite these naturally occurring lags among the senses, observers perceive intersensory synchrony for most multisensory events in the external world and not only for those at 15 m. Only in exceptional circumstances, such as the thunder that is heard *after* the lightning, is a single multisensory event perceived as being separated in time.

This raises the question of how temporal coherence is maintained, which is not only of interest for theoretical reasons but also of practical importance (e.g., video broadcasting or multimedia Internet, for which standards are required for allowable audio or video delays; Finger & Davis, 2001; Mortlock, Machin, McConnell, & Shepard, 1997; Rihs, 1995). In this overview, we describe four possible mechanisms of how intersensory synchrony might be maintained: (1) The brain might be insensitive for small lags, or it could just ignore them (a window of temporal integration); (2) the brain might be “intelligent” and bring deeply rooted knowledge about the external world into play that allows it to compensate for various external factors; (3) the brain might be flexible and shift its criterion about synchrony in an adaptive fashion; or (4) in order to reduce gaps, the brain might actively shift the time at which one information stream is perceived toward the other. Below, we address these notions more extensively. Note that none of these options mutually excludes another. There is, in fact, evidence for each of the four mechanisms, although not all of them are equally persuasive. We first spend a few words, however, on how perception of intersensory synchrony is usually measured and on factors that affect sensitivity for intersensory asynchrony.

How Synchrony Between Two Senses Is Measured

A classic way to measure intersensory synchrony is simply to ask participants to judge whether two events were simultaneous or successive. This is known as a *simultaneity judgment task* (SJ task). Typically, stimulus pairs are presented at various stimulus onset asynchronies (SOAs), and participants explicitly judge whether the stimuli were simultaneous or not. If the percentage of “simultaneous” responses is plotted as a function of the SOA, one usually obtains a bell-shaped Gaussian curve (see Figure 1). The peak of this curve is taken as the *point of subjective simultaneity* (PSS), in that it is assumed that, at this particular SOA, the information from the different modalities is perceived as being maximally simultaneous. The second measure that can be derived from this curve is its standard deviation (*SD*), which is reflected in the width of the curve. This width, or some measure derived from it, can, as a first approximation, be taken as the *window of temporal integration*, because it conceivably represents the range of SOAs at which the brain treats the two information streams as belonging to the same event.

Despite the fact that the interpretation of the SJ task seems rather straightforward, a major disadvantage of this task is that the size of the window depends heavily on subjective criterion settings. Thus, an observer who is quite willing to respond “simultaneous” will have a large win-

dow if compared with an observer with a more stringent criterion. Observers may also be inclined to assume that stimuli that naturally belong together are synchronous, and such a cognitive bias might then affect the size of the window. An experimenter may, for example, dub a male voice onto the video of a female face and ask participants whether the audio and video were synchronous (Vatakis & Spence, 2007). The gender-incongruent combinations may then more likely be judged as asynchronous, not because the face and voice were poorly matched in time, but because the gender mismatch is unnatural. Such a cognitive bias would then be reflected in the SJ task, and for that reason, the SJ task is quite often not particularly well-suited to properly measure the width of the window of temporal integration.

An alternative to the SJ task is the temporal order judgment task (TOJ task). In a TOJ task, stimuli are also presented at various SOAs, but, rather than judging whether stimuli were simultaneous or successive, observers judge which stimulus came first (or second). Participants in an audio-visual TOJ task thus respond “sound first” or “light first.” If the percentage of “light first” responses is plotted as a function of the SOA, one usually obtains an S-shaped logistic psychometric curve (see Figure 1). From this curve, one can again derive two measures: the 50% crossover point and the steepness of the curve at the 50% point. The 50% crossover point is the SOA at which observers presumably were maximally unsure (i.e., maximally simultaneous) about the temporal order, and is therefore taken as the PSS. The steepness at the crossover point reflects the observers’ sensitivity to temporal asynchronies. The steepness can also be expressed in terms of the *just noticeable difference* (JND—half the difference in SOA between the 25% and 75% point), and then it represents the smallest interval observers can reliably notice. A steep psychometric curve thus results in a small JND and implies a high temporal sensitivity (i.e., small asynchronies are still correctly perceived).

One might expect, as depicted in Figure 1, that the JND as measured in a TOJ task corresponds well with the width of the Gaussian curve obtained in an SJ task (i.e., the window of temporal integration), because temporal order should be difficult to judge if stimuli are perceived as simultaneous. A large temporal window should thus correspond with high JND values. The match between these two measures, however, is, in general, quite poor. It possibly reflects that judgments about simultaneity and temporal order are based on different information sources, because it is possible to perceive two stimuli as asynchronous, but not to know which came first (Hirsh & Fraisse, 1964; Mitrani, Shekerdjiski, & Yakimoff, 1986; Schneider & Bavelier, 2003; van Eijk, Kohlrausch, Juola, & van de Par, 2008, 2009; Zampini, Shore, & Spence, 2003a). Moreover, in the TOJ task, in which only temporal order responses can be given (“sound first” or “light first”), observers may be inclined to adopt the assumption that stimuli are never simultaneous; whereas in the SJ task, observers may be inclined to assume that stimuli belong together, only because the “synchronous” response category is available. Different criterion settings in the

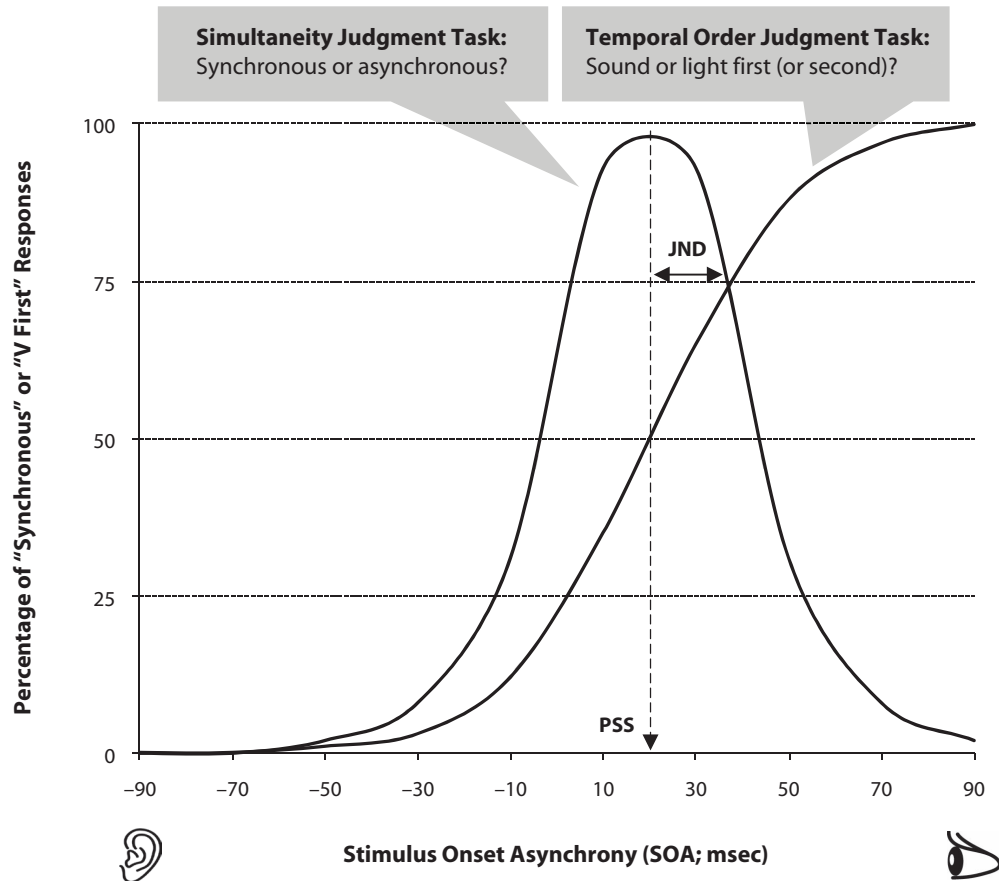


Figure 1. Hypothetical data of an audio–visual temporal order judgment (AV–TOJ) task (S-shaped curve) and a simultaneity judgment (SJ) task (bell-shaped curve). Auditory and visual stimuli are presented at various stimulus onset asynchronies (SOAs), ranging from clear auditory first (A first) to clear vision first (V first). In the TOJ task, participants judge which stimulus came first, whereas in the SJ task, they judge whether stimuli were simultaneous or successive. The point of subjective simultaneity (PSS) represents the interval at which the two stimuli are perceived as being maximally simultaneous (here at ~20 msec). In the SJ task, this is the SOA at which most synchronous responses were given—in the TOJ task, it is the point at which 50% of the responses are “V first” and 50% are “A first.” The just noticeable difference (JND) represents the smallest interval that observers can reliably notice (here around 20 msec). In the SJ task, this is the average interval (of A first and V first), at which the participant responds with 75% “synchronous” responses. In the TOJ task, it is the difference in SOA at the 25% and 75% point divided by two.

two tasks may then result in poor convergence of estimates of the temporal window of integration.

When Is *Simultaneous*?

The naive reader might think that stimuli from different modalities are perceived as being maximally simultaneous if they are presented the way nature does it, namely synchronous, and therefore at 0-msec SOA. Surprisingly, however, this is not usually the case. For audio–visual stimuli, the PSS is usually shifted toward a visual-lead stimulus, which means that simultaneity is maximally perceived if vision comes slightly before hearing (e.g., Kayser, Petkov, & Logothetis, 2008; Lewald & Guski, 2003; Lewkowicz, 1996; Slutsky & Recanzone, 2001; Zampini, Guest, Shore, & Spence, 2005; Zampini, Shore, & Spence, 2003a, 2005; see Figure 1). Moreover, the raw data of an SJ or TOJ task are usually not mirror symmetric around the PSS but are skewed toward more “simul-

taneous” responses on the “light first” side of the axis. This bias was already found in a classic study by Dixon and Spitz (1980). Here, participants monitored continuous videos consisting of an audio–visual speech stream or an object event consisting of a hammer hitting a peg. The videos started off in synchrony and were then gradually desynchronized in 51-msec steps, up to a maximum asynchrony of 500 msec. Observers were instructed to respond as soon as they noticed the audio–visual asynchrony. They were better at detecting the asynchrony if the sound preceded the video rather than if the video preceded the sound (131- vs. 258-msec thresholds for speech and 75- vs. 188-msec thresholds for the hammer, respectively). The PSS values also showed that simultaneity was perceived as maximal when the video preceded the audio by 120 msec for speech and by 103 msec for the hammer. Many other studies have reported this vision-first PSS (Dinnerstein & Zlotogura, 1968; Hirsh & Fraisse, 1964;

Jaśkowski, Jaroszyk, & Hojan-Jezierska, 1990; Keetels & Vroomen, 2005; Spence, Baddeley, Zampini, James, & Shore, 2003; Vatakis & Spence, 2006; Zampini et al., 2003a); however, some also reported opposite results (Bald, Berrien, Price, & Sprague, 1942; Rutschmann & Link, 1964; Teatini, Fernè, Verzella, & Berruecos, 1976; Vroomen, Keetels, de Gelder, & Bertelson, 2004). There have been many speculations about the underlying reason for this overall visual-lead asymmetry. One possibility, as reflected in simple reaction time (RT), is that auditory stimuli may be processed faster than visual stimuli, and so visual stimuli would need to be presented before sounds to compensate for their neural delay. Alternatively, observers might be tuned toward the natural situation, in which lights reach the sense organs before sounds do (King & Palmer, 1985). In both cases, there is then a preference that vision will have a head start over sound.

Besides the overall vision-first bias, however, there are many other reasons why the PSS can differ from 0-msec SOA. To point out a few, the PSS depends on, among others, stimulus intensity (more intense stimuli are processed faster or come to consciousness more quickly; see Allik & Kreegipuu, 1997; Jaśkowski, 1999; Jaśkowski & Verleger, 2000; Roufs, 1963; Sanford, 1971; Smith, 1933), the nature of the response that participants have to make (e.g., “Which stimulus came first?” vs. “Which stimulus came second?”; see Frey, 1990; Spence, Shore, & Klein, 2001), individual differences (Mollon & Perkins, 1996; Stone et al., 2001), and the modality or location to which attention is directed (Mattes & Ulrich, 1998; Schneider & Bavelier, 2003; Shore & Spence, 2005; Shore, Spence, & Klein, 2001; Stelmach & Herdman, 1991; Zampini, Shore, & Spence, 2005). Here, however, we are not intending to review all of these factors, but, rather, are focusing on how simultaneity is perceived despite lags.

1. The Window of Temporal Integration

The most straightforward reason why, despite gaps, information streams are perceived as being simultaneous is that the brain is simply not sensitive enough to notice the gap (see Figure 2, panel 1). Information streams below the gap threshold would thus be perceived to have occurred simultaneously. The JND, as measured in a TOJ or SJ task, might then be a good first approximation for the size of the window. Alternatively, it might also be that lags above the threshold can be noted (e.g., if attended) but are, under normal circumstances, ignored up to some point, so that the information streams are, despite the asynchrony, integrated, in order to enhance perception or memory. If this were the case, the JND, as measured in TOJ or SJ tasks (in which temporal lag is the relevant response dimension), would then be better interpreted as the *lower* limit of the temporal window for multisensory integration.

Assuming that JNDs are indeed relevant, the next question to ask would be what the actual values are for the respective modality pairings. Unfortunately, there is no straightforward answer, since JNDs have been shown to differ quite substantially across studies. In Hirsh and Sherrick's (1961) classic study on sensitivity for intersen-

sory synchrony, well-trained participants were presented with audio–visual, visual–tactile, and audio–tactile stimuli in a TOJ task and reported JNDs to be approximately 20 msec, irrespective of the modalities used. More recent studies, however, found substantially bigger JNDs and larger differences among the sensory modalities. For simple cross-modal stimuli, such as auditory beeps and visual flashes, JNDs have been reported on the order of approximately 25–50 msec (Keetels & Vroomen, 2005; Zampini, Guest, et al., 2005; Zampini et al., 2003a), but for audio–tactile pairs, Zampini, Brown, et al. (2005) obtained JNDs of about 80 msec, and for visual–tactile pairs, JNDs have been found on the order of 35–65 msec (Keetels & Vroomen, 2008b; Spence et al., 2001). In recent years, it has become increasingly clear why JNDs can differ substantially across studies. Below, we list several factors that are known to be of importance.

Spatial separation. Sensitivity for temporal order in TOJ tasks improves if the components of the cross-modal stimuli are spatially separated (i.e., lower JNDs; Bertelson & Aschersleben, 2003; Spence et al., 2003; Spence et al., 2001; Zampini, Guest, et al., 2005; Zampini, Shore, & Spence, 2003a, 2003b). Possibly, this occurs because stimuli from a single location may be more likely paired as a unitary event than are stimuli presented far apart, and this intermodal pairing would make the relative temporal order of the components lost. Another possibility, one with no temporal basis, is that observers may have extra spatial information on which to base their responses. Thus, observers may initially not know which modality had been presented first but still know on which side the first stimulus appeared, and they may then infer which modality had been presented first (Spence et al., 2003).

Stimulus complexity. The temporal order of brief simple stimuli, such as a flash and a beep, that have a rather sharp rise time of energy is much easier to perceive than are time-varying stimuli with shallower slopes (van der Burg, Cass, Olivers, Theeuwes, & Alais, 2009). This has important consequences, because the rise time of a sound is easily confounded with the distance of the sound (distant sounds have shallow rise times; Blauert, 1997), and rise time may also explain why determining temporal order for audio–visual speech can be notoriously difficult. In fact, the delays at which auditory and visual speech streams are perceived as synchronous are extremely wide (Conrey & Pisoni, 2006; Dixon & Spitz, 1980; Jones & Jarick, 2006; Stekelenburg & Vroomen, 2007; van Wassenhove, Grant, & Poeppel, 2007; Vatakis & Spence, 2006). For example, in van Wassenhove et al., observers in an SJ task judged whether congruent audio–visual speech stimuli and incongruent McGurk-like speech stimuli (McGurk & MacDonald, 1976) were synchronous. The authors found a temporal window of 203 msec for the congruent pairs (ranging from –76 msec, sound first, to +127 msec, vision first, with PSS at 26 msec, vision first) and a 159-msec window for the incongruent pairs (ranging from –40 to +119 msec, with PSS at 40 msec, vision first). Note that these windows are up to ~5 times bigger than those found for simple flashes and beeps (mostly below 50 msec; Hirsh & Sherrick, 1961; Keetels

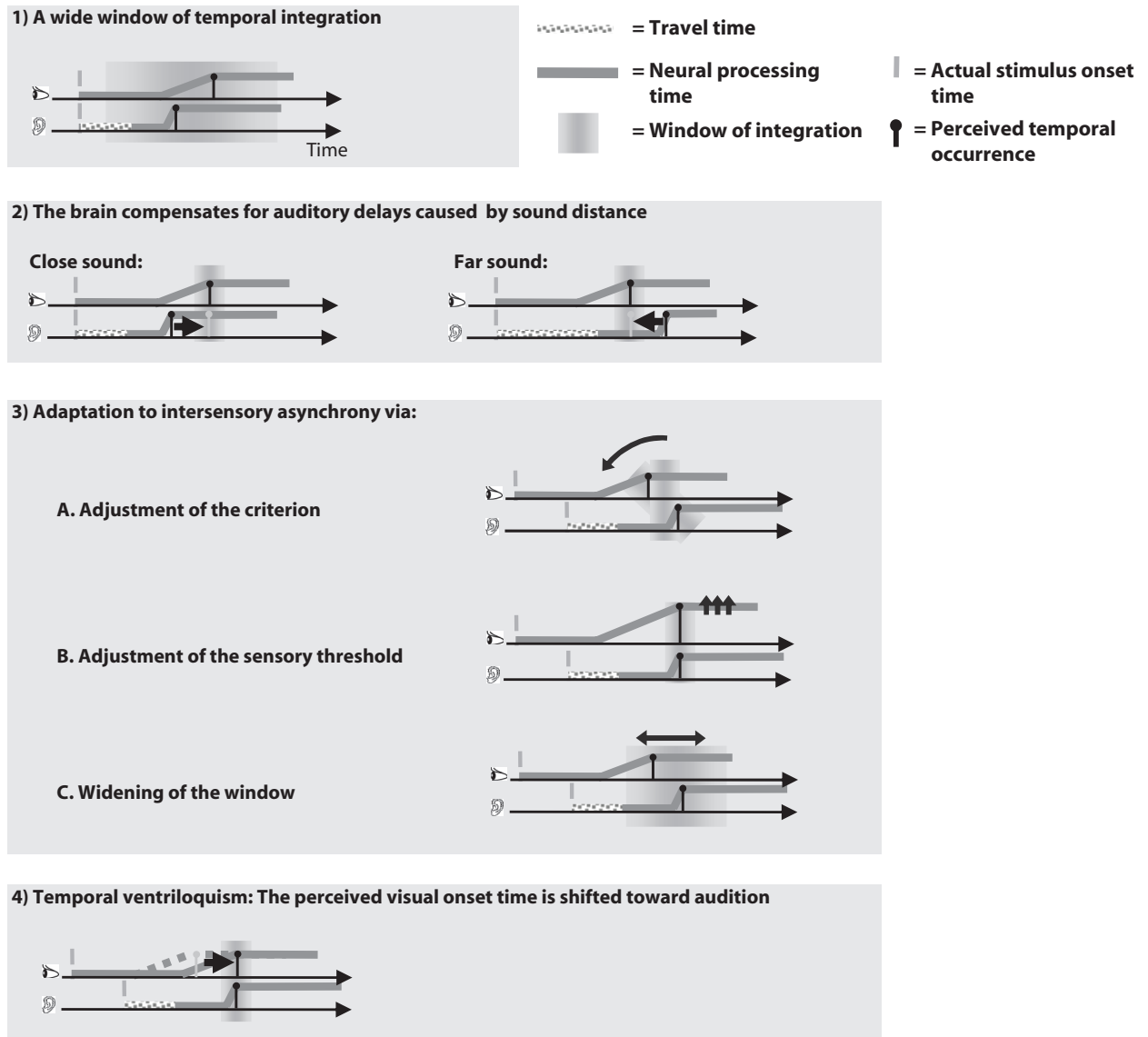


Figure 2. Synchrony can be perceived despite lags. How is this accomplished? Four possible mechanisms are depicted for audio–visual stimuli, such as flashes and beeps. Similar mechanisms might apply for other stimuli and other modality pairings. Time is represented on the *x*-axis, and accumulation of sensory evidence is represented on the *y*-axis. A stimulus is “time stamped” as soon as it surpasses a sensory threshold. Stimuli in audition and vision are perceived as being synchronous if they occur within a certain time window. (1) The brain might be insensitive for naturally occurring lags because the window of temporal integration is rather wide. (2) The brain might compensate for predictable variability (here, sound distance) by adjusting the perceived occurrence of a sound in accordance with sound travel time. (3) Temporal recalibration (three different mechanisms might underlie adaptation to asynchrony): (3A) A shift in criterion about synchrony for the adapted stimuli or modalities; (3B) a change in the threshold of sensory detection (when did the stimulus occur?) within one of the adapted modalities; (3C) a widening of the temporal window for the adapted stimuli or modalities. (4) Temporal ventriloquism: A visual event is actively shifted toward an auditory event.

& Vroomen, 2005; Zampini et al., 2003a; Zampini, Guest, et al., 2005). On the basis of these findings, some have concluded that speech is special (van Wassenhove et al., 2007; Vatakis, Ghazanfar, & Spence, 2008) or that, when stimulus complexity increases, sensitivity for temporal order deteriorates (Vatakis & Spence, 2006). It seems, however, that stimulus factors, such as rise time, need to be controlled more carefully before any sensible comparison can be made across audio–visual speech, complex stimuli, and simple combinations of flashes and beeps.

Frequency. For audio–visual stimuli that are repetitively presented in a stream, perception of synchrony breaks down if the temporal frequency is above ~4 Hz (Benjamins, van der Smagt, & Verstraten, 2008; Fujisaki & Nishida, 2005). Above this rate, observers are no longer able to discriminate whether the auditory and visual stimulus elements are synchronous, and the two modality streams are perceived as being segregated with no order between them (visual–tactile and auditory–tactile temporal resolution were ~4 Hz and ~10 Hz, respectively;

Fujisaki & Nishida, 2009). This limit at 4 Hz is rather low if compared with the unimodal perception of synchrony (e.g., deciding whether two flickering visual signals are in- or out-of-phase breaks down above ~25 Hz; Fujisaki & Nishida, 2005). Furthermore, it should be noted that the 4-Hz rate is also the approximate rate at which syllables are spoken in continuous speech, and one of the reasons why temporal order in continuous audio–visual speech might be difficult is that stimulus presentation is too fast. Note, however, that this cannot explain the whole picture, since temporal order in audio–visual speech remains difficult if individual syllables are used.

Predictive information. Sensitivity for temporal order improves if there is anticipatory information that predicts the onset of an audio–visual event. For many natural events, such as hand clapping, vision provides predictive information about when the sound is to occur, because there is visual anticipatory information about sound onset (i.e., the movement of the hands toward each other). Stimuli with predictive information allow observers to make a clear prediction about when auditory information should occur, and this may improve sensitivity for temporal order (Petrini, Russell, & Pollick, 2009; van Eijk, 2008, chapter 4; Vroomen & Stekelenburg, 2010). Multiple stimuli presented in a rhythmic sequence (although below 4 Hz) may also allow observers to make clear predictions about the expected stimulus onset, and rhythmic information may thus improve sensitivity for cross-modal temporal order.

Unity assumption. Observers may have difficulty judging temporal order for information streams that naturally belong together, as in the case of auditory and lipread speech. In the literature, this has been referred to as the *unity assumption* (Welch & Warren, 1980). Such stimuli may be strongly paired as a unitary event and are therefore perceived as being synchronous, despite lags. Evidence for this notion comes from Vatakis and Spence (2007), in which participants judged the temporal order of audio–visual speech that varied in gender and phonemic congruency. Speech tokens, such as /bi/, /o/, /pi/, and /a/, were produced by a male and a female voice and later dubbed onto the video of a male or female face articulating these tokens. In support of the unity assumption, results showed that sensitivity for temporal order was worse (larger JNDs) if the auditory and visual streams were gender congruent than if they were gender incongruent. More recently, however, Vatakis et al. (2008) qualified these findings and reported that this effect may be specific to human speech (because it is not observed in monkey calls or in human imitations of monkey calls). They examined the effect of congruency by using matching or mismatching types of monkey calls (“cooing” vs. “grunt” or threat calls), human speech, and human imitations of monkey calls. For audio–visual speech, sensitivity of temporal order was again worse for the gender-congruent than for the gender-incongruent trials, but there was no congruency effect for the monkey calls or for the human voice imitating the coos. Vatakis and Spence (2008) also found no congruency effect for audio–visual music and object events that either matched (e.g., the sight of a note being played on a piano together with the corresponding sound, the video of a bouncing ball with a

corresponding sound) or mismatched. At this point, it thus appears that congruency in intersensory pairing may play a role only in audio–visual speech.

To summarize, JNDs for intersensory temporal order can be as low as 20 msec, but the exact value depends on various other factors, such as spatial separation, frequency, and stimulus complexity.

As a first approximation, one might assume that the size of the window of temporal integration corresponds to the JND measured in TOJ or SJ tasks. Interestingly, however, this is not always the case, because there are also other, more *indirect* measures of the window of temporal integration that do not correspond with JNDs as measured in TOJ or SJ tasks. As an example, Munhall, Gribble, Sacco, and Ward (1996) demonstrated that exact temporal coincidence between the auditory and visual parts of audio–visual speech stimuli is not a very strict constraint on the McGurk effect. In the McGurk effect, a lipread speech token affects the phonetic content of a speech sound that is heard (McGurk & MacDonald, 1976). The effect was biggest when auditory vowels were synchronized with the original mouth movements (McGrath & Summerfield, 1985), but the effect survives, even if audition lagged vision by 180 msec (see also Soto-Faraco & Alsius, 2007, 2009; these studies show that participants can still perceive a McGurk effect when they can quite reliably perform TOJs at intervals above the JND). So, speech-sound identification can be influenced by lipread speech even if the two are perceived as being out of sync. Outside of the speech domain, similar findings have been reported. In a study by Shimojo et al. (2001), the role of temporal synchrony was examined using the *stream/bounce* illusion. Two visual targets that move across each other and are normally perceived as a streaming motion are perceived to bounce when a brief sound is presented at the moment that the visual targets coincide (Sekuler, Sekuler, & Lau, 1997). This phenomenon depends on the timing of the sound relative to the coincidence of the moving objects. Although it was demonstrated that a brief sound induced the visual bouncing percept most effectively when it was presented about 50 msec before the moving objects coincided, their data further showed a rather large temporal window of integration, inasmuch as intervals ranged from 250 msec before visual coincidence to 150 msec after coincidence (see also Bertelson & Aschersleben, 1998, for the effect of temporal asynchrony on spatial ventriloquism, or Shams, Kamitani, & Shimojo, 2002, for the illusory-flash effect).

These intersensory effects thus occur at asynchronies that are much larger than normally reported JNDs when directly asking for temporal order. Possibly, then, despite the fact that observers do notice small delays among the senses, the brain may still ignore these if it is of help to do so for other purposes, such as understanding speech (Soto-Faraco & Alsius, 2007, 2009).

2. Compensation for External Factors

The second explanation for how temporal coherence among the senses might be maintained is that the brain compensates for various predictable delays (Figure 2, panel 2). This controversial issue has received support

mainly from studies that examined whether observers take distance into account when judging audio–visual synchrony. As mentioned before, the relative slow transduction time of sounds through air causes natural differences in arrival time between sounds and lights. It implies that the farther away an audio–visual event is, the more the sound lags the visual stimulus. However, the brain might compensate for such a lag if distance is known. Some have reported that judgments about audio–visual synchrony were found to depend on perceived distance (Alais & Carlile, 2005; Engel & Dougherty, 1971; Kopinska & Harris, 2004; Sugita & Suzuki, 2003), but others failed to demonstrate these effects (Arnold, Johnston, & Nishida,

2005; Heron, Whitaker, McGraw, & Horoshenkov, 2007; Lewald & Guski, 2004; Stone et al., 2001). See Table 1 for an overview.

Sugita and Suzuki (2003) explored compensation for distance with an audio–visual TOJ task. The visual stimuli were delivered by light-emitting diodes (LEDs) at distances ranging from 1 to 50 m in free-field circumstances (compensated for by intensity, but not by size), and the sounds were delivered by headphones. PSS values were found to shift with visual stimulus distance. When the visual stimulus was 1 m away, the PSS was at about a ~5-msec sound delay, which increased when the LEDs were farther away. The increment was consistent with the

Table 1
Summary of Studies Examining Audio–Visual Synchrony

Author	Task and Stimuli	Results	Compensation for Sound Distance?	Criticism
Studies Showing an Effect				
Alais & Carlile (2005)	AV–TOJ task—Visual: blob on computer screen of fixed distance, Auditory: via a loud-speaker with distance simulated from 5–40 m.	PSS shifts with the apparent distance of the sound, in accordance with sound velocity through air. At far distance, a sound-late stimulus is perceived as synchronous.	Yes	Shift in PSS caused by poor sensitivity for distant sounds (low rise time). No attempt to equate visual and auditory distance. Adaptive staircase method to track the PSS might have led to recalibration effects.
Engel & Dougherty (1971)	AV–SJ task—Visual: white square on black background; Auditory: 0.5-msec click via collocated speaker. Observers seated at ~1, 12, 19, 27, 34 m.	PSSs remained constant up to ~20 m, beyond which the constancy begins to break down.	Yes, partly	Participants had to imagine AV colocalization, which might have led to strategic responses. Only 4 observers. Four out of the five distances were parallel with predicted distance for no compensation.
Kopinska & Harris (2004)	AV–TOJ task—Visual flash on PC monitor and tone burst from speaker. Observers seated at 1, 4, 8, 16, 24, and 32 m.	PSSs were not affected by distance.	Yes	Basically a null result. Distance was blocked over trials, possibly leading to adaptation (recalibration) or response equation.
Sugita & Suzuki (2003)	AV–TOJ task—Visual: LEDs at 1–50 m in free field; Auditory: via headphones (no attempt to simulate distance).	PSSs shifted in accordance with distance of visual stimulus. At far distance, sound late is perceived as synchronous.	Yes	Headphones are artificial. No attempt to equate the distance of the sound and the light. Participants had to imagine AV colocalization, which might have led to strategic responses.
Studies Not Showing an Effect				
Arnold, Johnston, & Nishida (2005)	Stream/bounce illusion and AV–TOJ task—Auditory: via loudspeakers or headphones. Observer distances: 1–15 m.	Optimal time to induce bounce illusion and PSSs shifted with sound distance if presented over loudspeakers but not if presented over headphones.	No compensation for distance if distance is real	
Heron, Whitaker, McGraw, & Horoshenkov (2007)	AV–TOJ task—Visual white circular disc on PC monitor and auditory noise click via speaker. Observer distances: 1, 5, 10, 20, 30, and 40 m.	PSS values shifted precisely with the variation in sound transmission time through air in both a corridor environment and a large reverberant environment. Compensation for distance was reported when participants first adapted to the stimuli.	No	Only 3 observers (2 authors).
Lewald & Guski (2004)	AV–TOJ task—A train of sounds and lights presented by collocated speakers/LEDs in open field. Observer distances: 1, 5, 10, 20, and 50 m.	PSS values shifted precisely with the variation in sound transmission time through air. At far distance, a sound-early stimulus is perceived as synchronous.	No	
Stone et al. (2001)	AV–SJ task—Observer distances: 0.5–3.5 m.	For 3 out of 5 participants, PSS shifted in correspondence with sound velocity.	No	Very small range of distances; effect observed in only 3 out of 5 participants.

velocity of sounds up to a viewing distance of about 10 m, after which it leveled off. This led the authors to conclude that lags between auditory and visual inputs are perceived as synchronous, not because the brain has a wide temporal window for audio–visual integration, but because the brain actively changes the temporal location of the window depending on the perceived distance of the source. Similar conclusions were reported by Alais and Carlile (2005), Engel and Dougherty (1971), and Kopinska and Harris (2004), although the latter two studies found that PSS values remained constant when observer–stimulus distance increased.

Others, however, failed to observe compensation for distance (Arnold et al., 2005; Heron et al., 2007; Lewald & Guski, 2004; Stone et al., 2001). For example, Lewald and Guski (2004) used a rather wide range of distances (1, 5, 10, 20, and 50 m), but their audio–visual stimuli (a sequence of five beeps/flashes) were delivered by colocalized speakers/LEDs placed in the open field. Note that, in this setup, there were no violations in the “naturalness” of the audio–visual stimuli, because they were physically colocalized. Here, the authors did not observe compensation for distance. Rather, their results showed that, when the physical observer–stimulus distance increased, the PSS shifted precisely with the variation in sound-transmission time through air. Note that this shift was precisely in the opposite direction from the PSS shifts that Sugita and Suzuki (2003) reported. Here, for audio–visual stimuli far away, sounds thus had to be presented earlier than for stimuli nearby, in order to be perceived as simultaneous, with no sign that the brain would compensate for sound-traveling time. The authors also suggested that the discrepancy between their findings and those that showed compensation for distance lies in the fact that the authors’ study *simulated* distance, whereas the others used the natural situation. Similar conclusions were reached by Arnold et al., who examined whether the stream/bounce illusion (Sekuler et al., 1997) varied with distance. They observed that the optimal time to produce a “bounce” percept varied with distance for sound presented over loudspeakers, but not for sound presented over headphones. This led the authors to conclude that there is no compensation for distance, if distance is real and presented over speakers rather than simulated and presented over headphones.

Apart from whether distance is real or simulated, other factors are conceivably at play. For example, Kopinska and Harris (2004) used real and colocalized audio–visual stimuli, but they nevertheless concluded that there was compensation for distance. In their study, the various distances were not randomized on a trial-by-trial basis, but were blocked. If distance is blocked, either adaptation to the additional sound lag may occur (i.e., recalibration, as described in the following section) or participants may equate response probabilities at the particular distance at which they are seated (for a similar interpretation, see Heron et al., 2007). Either way, the effect of distance on the PSS diminishes and no shift in the PSS is observed, misleading one to conclude that distance is compensated for. The same line of reasoning applies to Alais and Carlile’s (2005) study, in which an adaptive staircase method was

used to track the PSS. Another potentially relevant difference is the complexity of the stimuli. Unfortunately, none of the studies on distance compensation reported JNDs, but studies that concluded that there was compensation for distance used audio–visual stimuli (in particular, the farthest) whose temporal order was extremely difficult to judge. For the far stimuli, JNDs thus became poor, likely because far sounds have slow rise time, thereby shifting the PSS. This might easily lead one to falsely conclude that there is compensation for distance.

3. Temporal Recalibration

A third possibility of how the brain might deal with lags among the senses entails that the brain is flexible in adapting what it counts as synchronous (Figure 2, panel 3), a phenomenon known as *temporal recalibration*. Recalibration is a well-known phenomenon in the spatial domain, but it has been demonstrated only recently in the temporal domain (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vroomen et al., 2004). As for the spatial case, more than a century ago, Helmholtz (1867) showed that the visuo-motor system is remarkably flexible at adapting to shifts of the visual field induced by wedge prisms. If prism-wearing participants had to pick up a visually displaced object, they would quickly adapt to the new sensorimotor arrangement, and, even after only a few trials, small visual displacements might get unnoticed. *Recalibration* was the term used to explain this phenomenon. Recalibration is thought to be driven by a tendency of the brain to *minimize discrepancies among the senses* about objects or events that normally belong together. For the prism case, it is the position in which the hand is seen and felt. Nowadays, it is also known that the least reliable source is adjusted toward the more reliable one (Ernst & Banks, 2002; Ernst, Banks, & Bühlhoff, 2000; Ernst & Bühlhoff, 2004).

The first evidence of recalibration in the temporal domain came from two studies with very similar designs: an exposure–test paradigm. Both Fujisaki et al. (2004) and Vroomen et al. (2004) first exposed observers to a train of sounds and light flashes with a constant but small intersensory interval, and then tested them by using an audio–visual TOJ or SJ task. The idea was that observers would adapt to small audio–visual lags in such a way that the adapted lag eventually would be perceived as synchronous. So, after light-first exposure, light-first trials would be perceived as synchronous, and after sound-first exposure, a sound-first stimulus would be perceived as synchronous. Both studies indeed showed that the PSS was shifted in the direction of the exposure lag. For example, Vroomen and Keetels (2009) exposed participants for ~3 min to a sequence of sound bursts and light flashes with audio–visual lags of either ± 100 or ± 200 msec (sound first or light first). At test, the PSS was shifted an average of 27 and 18 msec (PSS difference between sound first and light first) for, respectively, the SJ and TOJ task. Fujisaki et al. used slightly bigger lags (± 235 msec, sound first or light first) and found somewhat bigger shifts in the PSS (59-msec shifts of the PSS in SJ and 51-msec shifts in TOJ), but data were comparable. Since then, many others have reported similar effects (Asakawa, Tanaka, & Imai,

2009; Di Luca, Machulla, & Ernst, 2009; Hanson, Heron, & Whitaker, 2008; Keetels & Vroomen, 2007, 2008b; Navarra, Hartcher-O'Brien, Piazza, & Spence, 2009; Navarra, Soto-Faraco, & Spence, 2007; Navarra et al., 2005; Stetson, Cui, Montague, & Eagleman, 2006; Sugita & Suzuki, 2003; Takahashi, Saiki, & Watanabe, 2008; Tanaka, Sakamoto, Tsumura, & Suzuki, 2009; Yamamoto, Miyazaki, Iwano, & Kitazawa, 2008).

The mechanism underlying temporal recalibration, however, remains elusive. It may be that there is a shift in the criterion for simultaneity in the adapted modalities (Figure 2, panel 3A). After exposure to light-first pairings, participants may thus change their criterion for audio-visual simultaneity in such a way that light-first stimuli are taken to be simultaneous; other modality pairings (e.g., vision-touch) would be unaffected, and the change in criterion should then not affect unimodal processing of visual and auditory stimuli presented in isolation. Another strong prediction is that stimuli that were synchronous before adaptation can become *asynchronous* after adaptation. The most dramatic case of this phenomenon can be found in motor-visual adaptation. In a study by Stetson et al., 2006, participants were asked to repeatedly tap their finger on a key, and, after each tap, a delayed flash was presented. If the visual flash occurred at an unexpectedly short delay after the tap (or was synchronous), it was perceived as occurring before the tap, an experience that runs against the law of causality.

It may also be the case that one modality (vision, audition, touch) is "shifted" toward the other, possibly because the sensory threshold for stimulus detection in one of the adapted modalities is changed (Figure 2, panel 3B). For example, as an attempt to perceive simultaneity during light-first exposure, participants might delay processing time in the visual modality by adopting a more stringent criterion for sensory detection of visual stimuli. After exposure to light-first audio-visual pairings, one might then expect slower processing times of visual stimuli in general, and other modality pairings that involve the visual modality, say vision-touch, would then also be affected.

Two strategies have been undertaken to explore the mechanism underlying temporal recalibration. The first is to examine whether temporal recalibration generalizes to other stimuli within the adapted modalities; the second is to examine whether temporal recalibration affects different modality pairings than the ones adapted. Fujisaki et al. (2004) demonstrated that the effect of adaptation to temporal misalignment was effective, even when the visual test stimulus was very different from the exposure situation. The authors exposed observers to asynchronous tone-flash stimulus pairs and later tested them on the "stream/bounce illusion" (Sekuler et al., 1997). Fujisaki et al. reported that the optimal delay for obtaining a bounce percept in the stream/bounce illusion was shifted in the same direction as the adapted lag. Furthermore, after exposure to a wall display, in which tones were timed with a ball bouncing off of the inner walls of a square, similar shifts in the PSS on the bounce percept were found (a ~45-msec difference when comparing the PSS of the -235-msec sound-first exposure with the +235-msec

vision-first exposure). Audio-visual temporal recalibration thus generalized well to other visual stimuli.

Navarra et al. (2005) and Vatakis et al. (2008) also tested generalization for audio-visual temporal recalibration using stimuli from different domains (speech/nonspeech). Their observers had to monitor a continuous speech stream for target words that were presented either in synchrony with the video of a speaker or with the audio stream lagging 300 msec behind. During the monitoring task, participants performed a TOJ (Navarra et al., 2005; Vatakis, Navarra, Soto-Faraco, & Spence, 2007) or SJ (Vatakis et al., 2008) task on simple flashes and white-noise bursts that were overlaid on the video. Their results showed that *sensitivity*, rather than a shift in the PSS, became worse if participants were exposed to desynchronized rather than synchronized audio-visual speech. Similar effects, larger JNDs, were found with music stimuli. This led the authors to conclude that the window of temporal integration was widened (Figure 2, panel 3C) due to asynchronous exposure (see also Navarra et al., 2007, for effects on JND after adaptation to asynchronous audio-tactile stimuli). The authors argued that this effect on the JND may reflect an initial stage of recalibration where a more lenient criterion is adopted about simultaneity. With prolonged exposure, participants may then shift the PSS. The authors also considered, but rejected, an alternative explanation: that participants became confused by the nonmatching exposure stimuli, which, as a result, may also have affected the JND rather than the PSS, since it added noise to the distribution.

A second way to study the underlying mechanisms of temporal recalibration is to examine whether temporal recalibration generalizes to different modality pairings. Hanson et al. (2008) explored whether a "supramodal" mechanism might be responsible for recalibration of multisensory timing. They examined whether adaptation to audio-visual, audio-tactile, and tactile-visual asynchronies (10-msec flashes, noise bursts, and taps on the left index finger) generalized across modalities. The data showed that a brief period of repeated exposure to ± 90 -msec asynchrony in any of these pairings resulted in shifts of about 70 msec of the PSS on subsequent TOJ tasks and that the size and nature of the shifts were very similar across all three pairings. This led them to conclude that there is a general mechanism. Opposite conclusions, however, were reached by Harrar and Harris (2005). They exposed participants for 5 min to audio-visual pairs with a fixed time lag (250 msec, light first) but did not obtain shifts in the PSSs for touch-light pairs. In an extension of this topic (Harrar & Harris, 2008), observers were exposed for 5 min to ~100-msec lags of light-first stimuli for the audio-visual case, and touch-first stimuli for the auditory-tactile and visual-tactile case. Participants were tested on each of these pairs before and after exposure. Shifts of the PSS in the predicted direction were found in the audio-visual exposure test stimuli but not for the other cases. Di Luca, Machulla, and Ernst (2007) also exposed participants to asynchronous audio-visual pairs (~200-msec lags of sound first and light first) and measured the PSS for audio-visual, audio-tactile, and

visual–tactile test stimuli. Besides obtaining a shift in the PSS for audio–visual pairs, they found that the effect generalized to audio–tactile but not to visual–tactile test pairs. This pattern made the authors conclude that adaptation resulted in a phenomenal shift of the auditory event (Di Luca et al., 2007).

Navarra et al. (2009) recently reported that the auditory rather than visual modality is more flexible. Participants were exposed to synchronous or asynchronous audio–visual stimuli (224 msec, V first, or 84 msec, A first, for 5 min of exposure), after which they performed a speeded reaction task on unimodal visual or auditory stimuli. In contrast with the idea that visual stimuli get adjusted in time to the relatively more accurate auditory stimuli (Hirsh & Sherrick, 1961; Shipley, 1964; Welch, 1999; Welch & Warren, 1980), their results seemed to show the opposite, namely that auditory rather than visual stimuli were shifted in time. The authors reported that simple RTs to sounds became approximately 20 msec faster after V-first exposure and about 20 msec slower after A-first exposure, whereas simple RTs for visual stimuli remained unchanged. They explained this finding by alluding to the idea that visual information can serve as the temporal anchor, because it offers a more nearly exact estimate of the time of occurrence of a distal event than does auditory information because the travel time of light is not perceptible. Further research is needed, however, to examine whether a change in simple RTs is truly reflecting a change in the timing of that event, because there is considerable evidence that the two do not always go hand in hand (e.g., RTs are more affected by variations in intensity than are TOJs; Jaśkowski & Verleger, 2000; Neumann & Niepel, 2004).

To summarize: As yet, there is no clear explanation for the mechanism underlying temporal recalibration, because there is discrepancy in the data regarding generalization across modalities. It seems safe to conclude that the audio–visual exposure–test situation is the one most reliable for obtaining a shift in the PSS. Arguably, audio–visual pairs are more flexible because the brain has to correct for timing differences between auditory and visual stimuli due to naturally occurring delays caused by distance. Tactile stimuli might be more rigid in time, because visual–tactile and audio–tactile events always occur at the body surface, so less compensation for latency differences might be required here. As mentioned above, a widening of the JND, rather than a shift in the PSS, has also been observed, which may reflect an initial stage of recalibration in which a more lenient criterion about simultaneity is adopted. The reliability of each modality on its own is also likely to play role. For visual stimuli, it is known that they are less reliable in time than are auditory or tactile stimuli (Fain, 2003), and, as a consequence, they may be more malleable (Ernst & Banks, 2002; Ernst et al., 2000; Ernst & Bühlhoff, 2004), but there is also evidence that, in fact, the auditory modality is shifted.

4. Temporal Ventriloquism

A fourth possibility of how the brain might deal with lags among the senses is that the perceived timing of a

stimulus in one modality may be actively shifted toward the other (Figure 2, panel 4), with perceptual consequences in the shifted modality. This phenomenon is named *temporal ventriloquism*, as an analogy with the spatial ventriloquist effect. For spatial ventriloquism, it has long been known that listeners who heard a sound while seeing a spatially displayed flash had the (false) impression that the sound originated from the flash. This phenomenon was named the *ventriloquist illusion*, because it was considered a stripped-down version of what the ventriloquist does on stage (see, for reviews, Bertelson, 1999; Vroomen & de Gelder, 2004a).

The temporal ventriloquist effect is analogous to the spatial variant, except that, here, a sound attracts a visual event in the time dimension rather than vision attracting sound in the spatial dimension. There are by now many demonstrations of this phenomenon, and below we describe several. They all show that small lags between vision and sound (or touch) are reduced, and, thus, may go unnoticed, because the timing of visual events is flexible and adjusts immediately (Fendrich & Corballis, 2001; Freeman & Driver, 2008; Getzmann, 2007; Keetels & Vroomen, 2008a; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Scheier, Nijhawan, & Shimojo, 1999; Vroomen & de Gelder, 2004b).

Scheier et al. (1999) were among the first to demonstrate temporal ventriloquism using a visual TOJ task. Observers were presented with two lights at various SOAs, one above and one below a fixation point, and their task was to judge whether the upper or lower light came first. To induce temporal ventriloquism, Scheier et al. added two sounds that could be presented either before the first and after the second light (condition AVVA) or between the two lights (condition VAAV). Note that they used a *visual* TOJ task and that sounds were task irrelevant. The results showed that observers were more sensitive (i.e., smaller JNDs) in the AVVA condition compared with the VAAV condition (JNDs were ~24 vs. ~39 msec, respectively). Presumably, the two sounds attracted the temporal occurrence of the two lights and thus effectively pulled the lights farther apart in the AVVA condition and closer together in the VAAV condition. In single-sound conditions, AVV and VVA, sensitivity was not different from a visual-only baseline, indicating that the effects were not due to the initial sound acting as a warning signal or to some cognitive factor related to the observer's awareness of the sounds.

Morein-Zamir et al. (2003) replicated these effects and further explored the sound–light intervals at which the effect occurred. Intervals of ~100 to ~600 msec were tested, and it was shown that, up to an interval of 200 msec, the second sound was mainly responsible for the temporal ventriloquist effect, whereas the interval of the first sound had little effect (Vroomen & Keetels, 2009). In another study (Keetels & Vroomen, 2008a), it was explored whether touch affects vision on the time dimension in a similar way as audition does (visual–tactile ventriloquism) and whether spatial disparity between the vibrator and lights modifies this effect. The results showed that tactile–visual stimuli behaved like audio–visual stimuli, in that tempo-

rally misaligned tactile stimuli captured the onsets of the lights and that spatial discordance between the stimuli did not harm this phenomenon.

Another demonstration of temporal ventriloquism came from a study by Vroomen and de Gelder (2004b). Here, temporal ventriloquism was demonstrated using the flash-lag effect (FLE). In the typical FLE (MacKay, 1958; Nijhawan, 1994, 1997, 2002), a flash appears to lag a moving visual stimulus, even though the stimuli are presented at the same location. To induce temporal ventriloquism, Vroomen and de Gelder added a single click slightly before, at, or after the flash (at intervals of 0, 33, 66, and 100 msec). The results showed that the sound attracted the temporal onset of the flash and shifted it on the order of ~5%. A sound occurring 100 msec before the flash thus made the flash appear ~5 msec earlier, and a sound 100 msec after the flash made the flash appear ~5 msec later. A synchronous sound also improved sensitivity on the visual task; JNDs were better if a sound was present rather than absent. Stekelenburg and Vroomen (2005) investigated the time course and the electrophysiological correlates of this flash-lag temporal ventriloquist effect by using event-related potentials (ERPs). In accordance with the behavioral findings, their results demonstrated that the amplitude of the visual N1 was systematically affected by the temporal interval between the visual target flash and the task-irrelevant sound in the FLE paradigm (MacKay, 1958; Nijhawan, 1994, 1997, 2002). If a sound was presented in synchrony with the flash, the N1 amplitude was larger than when the sound lagged behind the flash and smaller when it led the flash. However, no latency shifts were found. Yet, on the basis of the latency of the cross-modal effect (N1 at 190 msec) and its localization in the occipitoparietal cortex, this study confirmed the sensory nature of temporal ventriloquism.

In most of the studies examining temporal ventriloquism (visual TOJ, FLE, reporting clock position or motion direction), the timing of the visual stimulus is the task-relevant dimension. Recently, however, Vroomen and Keetels (2009) explored whether a temporally offset sound can improve identification of a visual stimulus in cases in which temporal order is not involved. In this study, it was examined whether four-dot masking is affected by temporal ventriloquism. In a typical four-dot masking paradigm, visual target identification is impaired when a briefly presented target is followed by a mask that consists of four dots that surround but do not touch the visual target (Enns, 2004; Enns & Di Lollo, 1997, 2000). The idea tested was whether a sound presented slightly before the target and slightly after the mask might lengthen the perceived interval between target and mask. By lengthening the perceived target-mask interval, there is more time for the target to consolidate, and, in turn, target identification should improve. Results were in line with this hypothesis, because a small release from four-dot masking was reported (1% improvement, which corresponds to a 4.4-msec increase of the target-mask interstimulus interval) if two sounds were presented at approximately 100-msec intervals before the target and after the mask, rather than if only a

single sound was presented before the target or a silent condition.

To summarize, there are many demonstrations that vision is flexible on the time dimension. The perceived timing of a visual event is attracted toward other events in audition and touch, provided that the lag between them is below ~200 msec. It implies that, in cross-modal SJ or TOJ tasks, where stimuli are typically presented at these short SOAs, there may always be a mutual attraction between the senses due to temporal ventriloquism. This attraction may partly explain why temporal order for multisensory stimuli is difficult if compared with unisensory performance. The deeper reason why there is this mutual attraction is as yet untested, but in our view, it serves to immediately reduce natural lags among the senses so that they get unnoticed, thus maintaining temporal coherence.

CONCLUSION

In recent years, a substantial amount of research has been devoted to understanding how the brain handles temporal asynchronies among the senses. Temporal lags below 20 msec are usually unnoticed, probably because of hard-wired limitations on the resolution power of the individual senses. Above this limit, delays do become noticeable, in particular if stimuli (1) have fast transient rise times, (2) are spatially separated, (3) have predictable onsets, and (4) are presented rhythmically at rates <4 Hz. It has been speculated that lags might become unnoticed because the brain is intelligent and compensates for predictable delays—in particular, delays caused by sound distance. This idea, however, is controversial and has not been demonstrated compellingly with natural stimuli (i.e., without headphones) and/or with good performance (e.g., low JNDs). Temporal lags among the senses may also go unnoticed, because there is mutual attraction among the senses. In part, it reflects a general tendency of the brain to reduce errors among the senses about information sources that normally yield converging data about the same event. This phenomenon has been demonstrated most clearly in temporal ventriloquism, for which the apparent time of a flash is shifted toward an abrupt sound or tap that shortly precedes or follows the flash. Exposure to intersensory delays can also result in adaptive shifts, a phenomenon called *temporal recalibration*. The mechanism underlying temporal recalibration, however, remains to be explored further. On the one hand, it may reflect that observers did adjust their criterion about what counts as synchronous for that particular modality pairing. Alternatively, it could also be that one modality (vision, audition, or touch) is shifted in time to compensate for the delay, possibly by adjusting the threshold for sensory detection in the adapted modality. The neural mechanisms that underlie these phenomena, including those among patients in whom there might be a breakdown, are of clear importance for future research.

AUTHOR NOTE

M.K. is supported by NWO-VENI Grant 451-08-020. An extensive review of this literature will appear in *Frontiers in the Neural Basis of*

Multisensory Processes (Micah M. Murray and Mark Thomas Wallace, Eds.). Correspondence relating to this article may be sent to J. Vroomen, Tilburg University, Department of Psychology, Warandelaan 2, 5037 AB, Tilburg, The Netherlands (e-mail: j.vroomen@uvt.nl).

REFERENCES

- ALAIS, D., & CARLILE, S. (2005). Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound. *Proceedings of the National Academy of Sciences*, **102**, 2244-2247.
- ALLIK, J., & KREEGIPUU, K. (1997). Multiple visual latency. *Perception*, **26** (ECVP Abstract Suppl.).
- ARNOLD, D. H., JOHNSTON, A., & NISHIDA, S. (2005). Timing sight and sound. *Vision Research*, **45**, 1275-1284.
- ASAKAWA, K., TANAKA, A., & IMAI, H. (2009, July). *Temporal recalibration in audio-visual speech integration using a simultaneity judgment task and the McGurk identification task*. Paper presented at the 31st Annual Meeting of the Cognitive Science Society, Amsterdam.
- BALD, L., BERRIEN, F. K., PRICE, J. B., & SPRAGUE, R. O. (1942). Errors in perceiving the temporal order of auditory and visual stimuli. *Journal of Applied Psychology*, **26**, 382-388.
- BEDFORD, F. L. (1989). Constraints on learning new mappings between perceptual dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 232-248.
- BENJAMINS, J. S., VAN DER SMAGT, M. J., & VERSTRATEN, F. A. J. (2008). Matching auditory and visual signals: Is sensory modality just another feature? *Perception*, **37**, 848-858.
- BERTELSON, P. (1999). Ventriloquism: A case of crossmodal perceptual grouping. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 347-362). Amsterdam: Elsevier, North-Holland.
- BERTELSON, P., & ASCHERSLEBEN, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, **5**, 482-489.
- BERTELSON, P., & ASCHERSLEBEN, G. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology*, **50**, 147-155.
- BLAUERT, J. (1997). *Spatial hearing: The psychophysics of human sound localization* (J. S. Allen, Trans.) (Rev. ed.). Cambridge, MA: MIT Press.
- CALVERT, G. [A.], SPENCE, C., & STEIN, B. E. (Eds.) (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.
- CONREY, B., & PISONI, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *Journal of the Acoustical Society of America*, **119**, 4065-4073.
- DI LUCA, M., MACHULLA, T.-K., & ERNST, M. O. (2007, July). *Perceived timing across modalities*. Paper presented at the International Intersensory Research Symposium 2007: Perception and Action, Sydney, Australia.
- DI LUCA, M., MACHULLA, T.-K., & ERNST, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, **9**, 1-16.
- DINNERSTEIN, A. J., & ZLOTOGURA, P. (1968). Intermodal perception of temporal order and motor skills: Effects of age. *Perceptual & Motor Skills*, **26**, 987-1000.
- DIXON, N. F., & SPITZ, L. (1980). The detection of auditory visual desynchrony. *Perception*, **9**, 719-721.
- ENGEL, G. R., & DOUGHERTY, W. G. (1971). Visual-auditory distance constancy. *Nature*, **234**, 308. doi:10.1038/234308a0
- ENNS, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research*, **44**, 1321-1331.
- ENNS, J. T., & DI LOLLO, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological Science*, **8**, 135-139.
- ENNS, J. T., & DI LOLLO, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, **4**, 345-352.
- ERNST, M. O., & BANKS, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, **415**, 429-433.
- ERNST, M. O., BANKS, M. S., & BÜLTHOFF, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, **3**, 69-73.
- ERNST, M. O., & BÜLTHOFF, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, **8**, 162-169.
- FAIN, G. L. (2003). *Sensory transduction*. Sunderland, MA: Sinauer.
- FENDRICH, R., & CORBALLIS, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, **63**, 719-725.
- FINGER, R., & DAVIS, A. W. (2001). *Measuring video quality in videoconferencing systems* (Tech. Rep. No. SN187-D). Los Gatos, CA: Pixel Instrument Corporation.
- FREEMAN, E., & DRIVER, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Current Biology*, **18**, 1262-1266.
- FREY, R. D. (1990). Selective attention, event perception and the criterion of acceptability principle: Evidence supporting and rejecting the doctrine of prior entry. *Human Movement Science*, **9**, 481-530.
- FUJISAKI, W., & NISHIDA, S. (2005). Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals. *Experimental Brain Research*, **166**, 455-464.
- FUJISAKI, W., & NISHIDA, S. (2009). Audio-tactile superiority over visuo-tactile and audio-visual combinations in the temporal resolution of synchrony perception. *Experimental Brain Research*, **198**, 245-259.
- FUJISAKI, W., SHIMOJO, S., KASHINO, M., & NISHIDA, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, **7**, 773-778.
- GETZMANN, S. (2007). The effect of brief auditory stimuli on visual apparent motion. *Perception*, **36**, 1089-1103.
- GIBSON, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- HANSON, J. V. M., HERON, J., & WHITAKER, D. (2008). Recalibration of perceived time across sensory modalities. *Experimental Brain Research*, **185**, 347-352.
- HARRAR, V., & HARRIS, L. R. (2005). Simultaneity constancy: Detecting events with touch and vision. *Experimental Brain Research*, **166**, 465-473.
- HARRAR, V., & HARRIS, L. R. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Experimental Brain Research*, **186**, 517-524.
- HELMHOLTZ, H. VON (1867). *Handbuch der physiologischen Optik*. Leipzig: Voss.
- HERON, J., WHITAKER, D., MCGRAW, P. V., & HOROSHENKOV, K. V. (2007). Adaptation minimizes distance-related audiovisual delays. *Journal of Vision*, **7**, 1-8.
- HIRSH, I. J., & FRAISSE, P. (1964). Simultanéité et succession de stimuli hétérogènes. *L'Année Psychologique*, **64**, 1-19.
- HIRSH, I. J., & SHERRICK, C. E., JR. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, **62**, 423-432.
- JAŚKOWSKI, P. (1999). Reaction time and temporal-order judgment as measures of perceptual latency: The problem of dissociations. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 265-282). Amsterdam: Elsevier.
- JAŚKOWSKI, P., JAROSZYK, F., & HOJAN-JEZIERSKA, D. (1990). Temporal-order judgments and reaction time for stimuli of different modalities. *Psychological Research*, **52**, 35-38.
- JAŚKOWSKI, P., & VERLEGER, R. (2000). Attentional bias toward low-intensity stimuli: An explanation for the intensity dissociation between reaction time and temporal order judgment? *Consciousness & Cognition*, **9**, 435-456.
- JONES, J. A., & JARICK, M. (2006). Multisensory integration of speech signals: The relationship between space and time. *Experimental Brain Research*, **174**, 588-594.
- KAYSER, C., PETKOV, C. I., & LOGOTHETIS, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex*, **18**, 1560-1574.
- KEETELS, M., & VROOMEN, J. (2005). The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Experimental Brain Research*, **167**, 635-640.
- KEETELS, M., & VROOMEN, J. (2007). No effect of auditory-visual spatial disparity on temporal recalibration. *Experimental Brain Research*, **182**, 559-565.
- KEETELS, M., & VROOMEN, J. (2008a). Tactile-visual temporal ventriloquism: No effect of spatial disparity. *Perception & Psychophysics*, **70**, 765-771. doi:10.3758/PP.70.5.765
- KEETELS, M., & VROOMEN, J. (2008b). Temporal recalibration to tactile-visual asynchronous stimuli. *Neuroscience Letters*, **430**, 130-134.
- KING, A. J., & PALMER, A. R. (1985). Integration of visual and auditory

- information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research*, **60**, 492-500.
- KOPINSKA, A., & HARRIS, L. R. (2004). Simultaneity constancy. *Perception*, **33**, 1049-1060.
- LEWALD, J., & GUSKI, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, **16**, 468-478.
- LEWALD, J., & GUSKI, R. (2004). Auditory-visual temporal integration as a function of distance: No compensation for sound-transmission time in human perception. *Neuroscience Letters*, **357**, 119-122.
- LEWKOWICZ, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 1094-1106.
- MACEFIELD, G., GANDEVIA, S. C., & BURKE, D. (1989). Conduction velocities of muscle and cutaneous afferents in the upper and lower limbs of human subjects. *Brain*, **112**, 1519-1532.
- MACKEY, D. M. (1958). Perceptual stability of a stroboscopically lit visual field containing self-luminous objects. *Nature*, **181**, 507-508. doi:10.1038/181507a0
- MATTES, S., & ULRICH, R. (1998). Directed attention prolongs the perceived duration of a brief stimulus. *Perception & Psychophysics*, **60**, 1305-1317.
- MCGRATH, M., & SUMMERFIELD, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America*, **77**, 678-685.
- MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748.
- MITRANI, L., SHEKERDJISKI, S., & YAKIMOFF, N. (1986). Mechanisms and asymmetries in visual perception of simultaneity and temporal order. *Biological Cybernetics*, **54**, 159-165.
- MOLLON, J. D., & PERKINS, A. J. (1996). Errors of judgment at Greenwich in 1796. *Nature*, **380**, 101-102.
- MOREIN-ZAMIR, S., SOTO-FARACO, S., & KINGSTONE, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, **17**, 154-163.
- MORTLOCK, A. N., MACHIN, D., MCCONNELL, S., & SHEPPARD, P. (1997). Virtual conferencing. *BT Technology Journal*, **15**, 120-129.
- MUNHALL, K. G., GRIBBLE, P., SACCO, L., & WARD, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, **58**, 351-362.
- NAVARRA, J., HARTCHER-O'BRIEN, J., PIAZZA, E., & SPENCE, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proceedings of the National Academy of Sciences*, **106**, 9169-9173.
- NAVARRA, J., SOTO-FARACO, S., & SPENCE, C. (2007). Adaptation to audiotactile asynchrony. *Neuroscience Letters*, **413**, 72-76.
- NAVARRA, J., VATAKIS, A., ZAMPINI, M., SOTO-FARACO, S., HUMPHREYS, W., & SPENCE, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research*, **25**, 499-507.
- NEUMANN, O., & NIEPEL, M. (2004). Timing of "perception" and perception of "time." In C. Kaernbach, E. Schröger, & H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition* (pp. 245-269). Mahwah, NJ: Erlbaum.
- NIJHAWAN, R. (1994). Motion extrapolation in catching. *Nature*, **370**, 256-257.
- NIJHAWAN, R. (1997). Visual decomposition of colour through motion extrapolation. *Nature*, **386**, 66-69.
- NIJHAWAN, R. (2002). Neural delays, visual motion and the flash-lag effect. *Trends in Cognitive Sciences*, **6**, 387-393.
- PETRINI, K., RUSSELL, M., & POLLOCK, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition*, **110**, 432-439.
- PÖPPEL, E. (1985). *Grenzen des Bewußtseins*. Stuttgart: Deutsche Verlags-Anstalt. [Translated as *Mindworks: Time and conscious experience*. New York: Harcourt Brace Jovanovich, 1988.]
- PÖPPEL, E., SCHILL, K., & VON STEINBÜCHEL, N. (1990). Sensory integration within temporally neutral systems states: A hypothesis. *Naturwissenschaften*, **77**, 89-91. doi:10.1007/BF01131783
- RADEAU, M. (1994). Auditory-visual spatial interaction and modularity. *Cahiers de Psychologie Cognitive*, **13**, 3-51.
- RIHS, S. (1995). The influence of audio on perceived picture quality and subjective audio-visual delay tolerance. In R. Hamberg & H. de Ridder (Eds.), *Advanced methods for the evaluation of television picture quality: Proceedings of the MOSAIC workshop* (pp. 133-137). Eindhoven: Institute for Perception Research.
- ROUFS, J. A. J. (1963). Perception lag as a function of stimulus luminance. *Vision Research*, **3**, 81-91.
- RUTSCHMANN, J., & LINK, R. (1964). Perception of temporal order of stimuli differing in sense mode and simple reaction time. *Perceptual & Motor Skills*, **18**, 345-352.
- SANFORD, A. J. (1971). Effects of changes in the intensity of white noise on simultaneity judgments and simple reaction time. *Quarterly Journal of Experimental Psychology*, **23**, 296-303.
- SCHEIER, C. R., NIJHAWAN, R., & SHIMOJO, S. (1999). Sound alters visual temporal resolution. *Investigative Ophthalmology & Visual Science*, **40**, 4169.
- SCHNEIDER, K. A., & BAVELIER, D. (2003). Components of visual prior entry. *Cognitive Psychology*, **47**, 333-366.
- SEKULER, R., SEKULER, A. B., & LAU, R. (1997). Sound alters visual motion perception. *Nature*, **385**, 308. doi:10.1038/385308a0
- SHAMS, L., KAMITANI, Y., & SHIMOJO, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, **14**, 147-152.
- SHIMOJO, S., SCHEIER, C., NIJHAWAN, R., SHAMS, L., KAMITANI, Y., & WATANABE, K. (2001). Beyond perceptual modality: Auditory effects on visual perception. *Acoustical Science & Technology*, **22**, 61-67.
- SHIPLEY, T. (1964). Auditory flutter-driving of visual flicker. *Science*, **145**, 1328-1330.
- SHORE, D. I., & SPENCE, C. (2005). Prior entry. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 89-95). Amsterdam: Elsevier, North-Holland.
- SHORE, D. I., SPENCE, C., & KLEIN, R. M. (2001). Visual prior entry. *Psychological Science*, **12**, 205-212.
- SLUTSKY, D. A., & RECANZONE, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *NeuroReport*, **12**, 7-10.
- SMITH, W. F. (1933). The relative quickness of visual and auditory perception. *Journal of Experimental Psychology*, **16**, 239-257.
- SOTO-FARACO, S., & ALSIUS, A. (2007). Conscious access to the unisensory components of a cross-modal illusion. *NeuroReport*, **18**, 347-350.
- SOTO-FARACO, S., & ALSIUS, A. (2009). Deconstructing the McGurk-MacDonald illusion. *Journal of Experimental Psychology: Human Perception & Performance*, **35**, 580-587.
- SPENCE, C. (2007). Audiovisual multisensory integration. *Acoustic Science & Technology*, **28**, 61-70.
- SPENCE, C., BADDELEY, R., ZAMPINI, M., JAMES, R., & SHORE, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. *Perception & Psychophysics*, **65**, 318-328.
- SPENCE, C., SHORE, D. I., & KLEIN, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General*, **130**, 799-832.
- STEIN, B. E., & MEREDITH, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press, Bradford Books.
- STEKELBURG, J. J., & VROOMEN, J. (2005). An event-related potential investigation of the time-course of temporal ventriloquism. *NeuroReport*, **16**, 641-644.
- STEKELBURG, J. J., & VROOMEN, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, **19**, 1964-1973.
- STELMACH, L. B., & HERDMAN, C. M. (1991). Directed attention and perception of temporal order. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 539-550.
- STETSON, C., CUI, X., MONTAGUE, P. R., & EAGLEMAN, D. M. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron*, **51**, 651-659.
- STONE, J. V., HUNKIN, N. M., PORRILL, J., WOOD, R., KEELER, V., BEANLAND, M., ET AL. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society B*, **268**, 31-38.
- SUGITA, Y., & SUZUKI, Y. (2003). Implicit estimation of sound-arrival time. *Nature*, **421**, 911.
- SUMBY, W. H., & POLLACK, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, **26**, 212-215.
- SUMMERFIELD, A. Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3-51). London: Erlbaum.

- TAKAHASHI, K., SAIKI, J., & WATANABE, K. (2008). Realignment of temporal simultaneity between vision and touch. *NeuroReport*, **19**, 319-322.
- TANAKA, A., SAKAMOTO, S., TSUMURA, K., & SUZUKI, Y. (2009). Visual speech improves the intelligibility of time-expanded auditory speech. *NeuroReport*, **20**, 473-477.
- TEATINI, G., FERNÉ, M., VERZELLA, F., & BERRUECOS, P., JR. (1976). Perception of temporal order: Visual and auditory stimuli. *Giornale Italiano di Psicologia*, **3**, 157-164.
- VAN DER BURG, E., CASS, J., OLIVERS, C. N. L., THEEUWES, J., & ALAIS, D. (2009). Efficient visual search from nonspatial auditory cues requires more than temporal synchrony. In E. van der Burg (Ed.), *Temporal multisensory processing and its effects on attention* (pp. 63-84). Amsterdam: Vrije Universiteit.
- VAN EIJK, R. L. J. (2008). *Audio-visual synchrony perception*. Unpublished doctoral dissertation, Technische Universiteit Eindhoven, Eindhoven.
- VAN EIJK, R. L. J., KOHLRAUSCH, A., JUOLA, J. F., & VAN DE PAR, S. (2009). Efficient visual search from nonspatial auditory cues requires more than temporal synchrony. In E. van der Burg (Ed.), *Temporal multisensory processing and its effects on attention* (pp. 63-84). Amsterdam: Vrije Universiteit.
- VAN EIJK, R. L. J., KOHLRAUSCH, A., JUOLA, J. F., & VAN DE PAR, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & Psychophysics*, **70**, 955-968.
- VAN EIJK, R. L. J., KOHLRAUSCH, A., JUOLA, J. F., & VAN DE PAR, S. (2008). Temporal interval discrimination thresholds depend on perceived synchrony for audio-visual stimulus pairs. *Journal of Experimental Psychology: Human Perception & Performance*, **35**, 1254-1263.
- VAN WASSENHOVE, V., GRANT, K. W., & POEPEL, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, **45**, 598-607.
- VATAKIS, A., GHAZANFAR, A. A., & SPENCE, C. (2008). Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *Journal of Vision*, **8**, 1-11.
- VATAKIS, A., NAVARRA, J., SOTO-FARACO, S., & SPENCE, C. (2007). Temporal recalibration during asynchronous audiovisual speech perception. *Experimental Brain Research*, **181**, 173-181.
- VATAKIS, A., & SPENCE, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research*, **1111**, 134-142.
- VATAKIS, A., & SPENCE, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, **69**, 744-756.
- VATAKIS, A., & SPENCE, C. (2008). Evaluating the influence of the "unity assumption" on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica*, **127**, 12-23.
- VROOMEN, J., & DE GELDER, B. (2004a). Perceptual effects of cross-modal stimulation: Ventriloquism and the freezing phenomenon. In G. [A.] Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 141-150). Cambridge, MA: MIT Press.
- VROOMEN, J., & DE GELDER, B. (2004b). Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception & Performance*, **30**, 513-518.
- VROOMEN, J., & KEETELS, M. (2009). Sounds change four-dot masking. *Acta Psychologica*, **130**, 58-63.
- VROOMEN, J., KEETELS, M., DE GELDER, B., & BERTELSON, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, **22**, 32-35.
- VROOMEN, J., & STEKELENBURG, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audio-visual stimuli. *Journal of Cognitive Neuroscience*, **22**, 1583-1596. doi:10.1162/jocn.2009.21308
- WELCH, R. B. (1999). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 371-387). Amsterdam: Elsevier.
- WELCH, R. B., & WARREN, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, **88**, 638-667.
- YAMAMOTO, S., MIYAZAKI, M., IWANO, T., & KITAZAWA, S. (2008, July). *Bayesian calibration of simultaneity in audiovisual temporal order judgment*. Poster presented at the 9th International Multisensory Research Forum, Hamburg.
- ZAMPINI, M., BROWN, T., SHORE, D. I., MARAVITA, A., RÖDER, B., & SPENCE, C. (2005). Audiotactile temporal order judgments. *Acta Psychologica*, **118**, 277-291.
- ZAMPINI, M., GUEST, S., SHORE, D. I., & SPENCE, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics*, **67**, 531-544.
- ZAMPINI, M., SHORE, D. I., & SPENCE, C. (2003a). Audiovisual temporal order judgments. *Experimental Brain Research*, **152**, 198-210.
- ZAMPINI, M., SHORE, D. I., & SPENCE, C. (2003b). Multisensory temporal order judgments: The role of hemispheric redundancy. *International Journal of Psychophysiology*, **50**, 165-180.
- ZAMPINI, M., SHORE, D. I., & SPENCE, C. (2005). Audiovisual prior entry. *Neuroscience Letters*, **381**, 217-222.

(Manuscript received October 2, 2009;
revision accepted for publication January 8, 2010.)